1981

Evaluation of Sampling Technique of Ad-Valorem Survey

Ministry of Treasury and Economics

Central Statistical Services

October, 1981

99jw

CONTENT

1. INTRODUCTION

In August 1981, Revenue and Operations Research Branch asked Central Statistical Services to evaluate the sampling methodology used in gathering samples, and to analyse the sample data already collected for the Treasury Budget Proposal regarding ad-valorem taxation of Gasoline and Tobacco products. In the Bills 72, 73 and 76, it was established in the legislation implementing the Treasurer's budget proposals regarding ad-valorem taxation of Motor Fuels and Tobacco products, that the taxable price to which the tax rates are to be applied, would be based on the median price obtained by the Minister of Revenue from such periodic sampling of retail price as he considers appropriate.

Staff in the Ministry of Revenue, with the concurrence of Treasury, had established that the sample, each quarter, would be drawn from the area of Ontario bounded by Oshawa on the east, Barrie on the north, Kitchener on the west and Niagara Falls on the south. The collected data from the sample surveys was given to Central Statistical Services in late September for analysis.

2. PURPOSE

There are two basic objectives of this study:

i) To evaluate the sampling technique used by the Ministry of Revenue to collect sample data;

ii) Based on the analysis of survey data collected by the Ministry of Revenue, recommend improvements in sampling methodology and estimation procedures of the taxable price of Motor Fuels and Tobacco products.

3. SUMMARY AND CONCLUSIONS

- Presently, the Ministry of Revenue is collecting data by judgement sample. Uniformed fuel tax inspectors collect data following the predetermined routes in the area specified in this study. Judgement sampling permits the sampler to select any sample and the probability of this selection is unknown. There is no assurance that the data collected represents the target population. In addition, certain errors may be introduced which cannot be easily measured, e.g. bias. In this study, we do not advise continuation of judgement sampling; we advise the selection of a probability sample determined by statistical sampling method, lest the methodology be indefensible.

- Presently, the selected sample sizes (500 to 600) for the survey are too large; we advise that sample size be selected according to accuracy and confidence level needed in the results. This will result in cost savings.

- Presently, only the median is calculated; we advise that all basic statistics should be determined and sampling errors should be determined for the sample data. After analysis of data, we have found that there is no significant difference among the arithmetic mean, median and mode. Since the mean and median are the same, the arithmetic mean can be used to discuss errors and confidence

limits of the population average.

- Presently, no error analysis is performed to indicate confidence in the results. It is necessary to estimate sampling errors and confidence limits in the survey data. (This allows the Ministry of Revenue to answer the fundamental question: How good is the estimate?)
- Since the variability in the sample data is small, the survey data indicates the following:
  - the data observations of all variables (Motor Fuels and Tobacco products) are quite homogeneous;
  - the geographic variation is very small between Metro and outside Metro;
  - there is no significant difference between estimated prices in Metro and outside Metro;
  - the arithmetic mean of various routes shows no significant difference between them.

4. RECOMMENDATIONS

The following recommendations are made to improve the survey methodology and estimation procedures:

i) It is highly recommended that recognized sampling procedures be used for sample survey of Motor Fuels and Tobacco products, i.e., systematic random sampling, stratified or cluster sampling with random selection.

ii) Sample size for the sample survey should be selected to reflect accuracy and confidence level required in the final results.

iii) It is highly recommended that error analysis should be performed and confidence limits should be determined for the estimation of population average.

5. DATA ANALYSIS OF SAMPLE DATA

The sample data (first quarter) collected by the Ministry of Revenue was analyzed in detail:

A. Estimation of Average Prices

In Tables 1 and 2, the median, mode and arithmetic mean are estimated and compared. The sample data indicates that there is no significant difference between the arithmetic mean, median and mode. It simply means that the prices of various grades are normally distributed. In general, the prices of the Motor Fuel products will be normally distributed, hence the arithmetic mean could be used to calculate errors and confidence limits. The median is a poor estimator of average, since basic statistical measurements cannot be calculated with the median.

B. Statistical Parameters of Sample Data

Table 3 provides estimates of statistical parameters which are needed to estimate the accuracy and precision of estimates of the average price of various grades of gasoline.

C. Confidence Limits

Table 4 determines population mean of various grades of Motor Fuel and provides confidence limits. It should be noted that the estimates at 95% confidence level are very precise. This is in part due to the large sample size and partly due to a good judgement sample.

D. Optimum Sample Size

Table 5 provides estimation of optimum sample sizes at various levels of tolerated errors, e.g., if accuracy to half

a cent is required for the estimation of average price with 95% confidence level, sample sizes in Column 6 will provide the required results. Similarly, Column 9 will provide sample sizes to achieve accuracy up to 1/5 of a cent.

E. Analysis of Tobacco and Cigarette Data

Tables 6, 7, 8, 9 and 10 provide information on estimation of average price, statistical parameters of sample, confidence limits and optimum sample size as described above in A, B, C and D.

F. Metro Sample vs Outside Metro Sample

Data collected in Metro and outside Metro was analyzed for comparison purposes. Table 11 includes statistical parameters for all fuel variables, tobacco and cigarettes. As can be seen, there is no significant difference in average prices between Metro and outside Metro.

G. Comparison of average prices of various routes

Table 12 provides the arithmetic mean by route of various grades of fuel. There is no significant difference in average prices of various grades of fuel between routes. The average price of fuel in various routes is almost the same as the mean price of the population mean. Statistically there is no difference in these averages.

6. TABLES

Table 1

Three Different Average Prices from Sample

Grade of Motive Fuel	Arithmetic Mean	The Median	The Mode
Regular Leaded	27.7¢	27.2¢	27.2¢
Regular Unleaded	29.9	29.3	29.2
Premium Leaded	31.2	31.5	29.4
Premium Unleaded	31.1	30.4	30.2
Diesel	26.3	26.0	25.9

Table 2

Comparison Among Three Different Average Prices

Grade	Mean-Median	Mean-Mode	Median-Mode
Regular Leaded	0.5¢	0.5¢	0
Regular Unleaded	0.6	0.7	0.1
Premium Leaded	-0.25	1.8	2.1
Premium Unleaded	0.7	0.9	0.2
Diesel	0.3	0.4	0.1

Table 3

Estimation of Statistical Parameters from Sample Data

Grade of Motive Fuel	Sample n	Mean $\bar{x}$	Variance S_x^2	Standard Deviation $S_{\bar{x}}$	Coefficient of Variation $S_{\bar{x}} / \bar{x}$	Standard Error $S_{\bar{x}} / \sqrt{n}$
	(Unit)	(Cent)	(Cent)	(Cent)	(Ratio)	(Cent)
Regular Leaded	457	27.7	2.07848	1.44187	5.20418	.06744
Regular Unleaded	452	29.9	2.36978	1.53941	5.15272	.07240
Premium Leaded	32	31.2	3.58065	1.89226	6.06008	.33450
Premium Unleaded	379	31.1	2.66794	1.63338	5.25119	.09918
Diesel	98	26.3	2.45773	1.56772	5.95443	.15836

Table 4

Estimation of Population Mean at 95% Confidence Limits

Grade of Motive Fuel	Sample Mean $\bar{x}$	Population Mean $\bar{X} = \bar{x} \pm t(\frac{S_{\bar{x}}}{\sqrt{n}})$	Estimated Population Mean $\bar{X}$
	(Cent)	(Cent)	(Cent)
Regular Leaded	27.7	$27.7 \pm 1.96(.06744)$	$27.6 \leq \bar{X} \leq 27.8$
Regular Unleaded	29.9	$29.9 \pm 1.96(.07240)$	$29.8 \leq \bar{X} \leq 30.0$
Premium Leaded	31.2	$31.2 \pm 1.96(.33450)$	$30.5 \leq \bar{X} \leq 31.9$
Premium Unleaded	31.1	$31.1 \pm 1.96(.09918)$	$30.9 \leq \bar{X} \leq 31.3$
Diesel	26.3	$26.3 \pm 1.96(.15836)$	$26.0 \leq \bar{X} \leq 26.6$

$$\bar{X} = \bar{x} \pm t(\frac{S_{\bar{x}}}{\sqrt{n}})$$

Where t = 1.96 at 95% confidence probability

Table 5

Estimation of Optimum Sample Sizes at Various Levels of Tolerated Errors

Grade of Motive Fuel	Optimum Sample Size n_i Required									
	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}
	1¢	.9	.8	.7	.6	.5	.4	.3	.2	.1
Regular Leaded	8	10	12	16	22	32	50	89	200	799
Regular Unleaded	9	11	14	19	25	36	57	101	228	910
Premium Leaded	14	17	21	28	38	55	86	153	344	1,379
Premium Unleaded	10	13	16	21	28	41	64	114	256	1,025
Diesel	9	12	15	19	26	38	59	105	236	944

$$n_i = \frac{t^2 S^2_{\bar{x}}}{e^2} = \left(\frac{t\, S_{\bar{x}}}{e}\right)^2, \quad i = 1,2, \ldots. 10$$

Where n_i = Optimum Sample Size required

t = 1.96 at 95% Confidence Level

$S^2_{\bar{x}}$ = Variance and $S_{\bar{x}}$ = Standard Deviation,

and

e = tolerated error in terms of cent per litre

1, .9, .8, .7, .6, .5, .4, .3, .2, .1.

Table 6

Three Different Average Prices from Sample

Tobacco and Cigarettes	Arithmetic Mean	The Median	The Mode
Tobacco	$1.41	$1.43	$1.45
Cigarettes - All sizes	1.07	1.05	1.05
Regular size	1.06	1.05	1.05
King size	1.07	1.05	1.05
100's	1.08	1.05	1.05

Table 7

Comparison Among Three Different Average Prices

Tobacco and Cigarettes	Mean-Median	Mean-Mode	Median-Mode
Tobacco	-$0.02	-$0.04	-$0.02
Cigarettes - All sizes	.02	.02	0
Regular size	.01	.01	0
King size	.02	.02	0
100's	.03	.03	0

Table 8

Estimation of Statistical Parameters from Sample Data

Tobacco and Cigarettes	Sample n	Mean $\bar{x}$	Variance $S^2_{\bar{x}}$	Standard Deviation $S_{\bar{x}}$	Coefficient of Variation $S_{\bar{x}} / \bar{x}$	Standard Error $S_{\bar{x}}/\sqrt{n}$
	Unit	$	$	$	Ratio	$
Tobacco	556	1.41	.0208921	.144541	10.2338	.0061299
Cigarettes - All sizes	1,684	1.07	.00464539	.0681571	6.37134	.00166089
Regular size	592	1.06	.0031255	.0559062	5.26544	.00229773
King size	589	1.07	.00503134	.070932	6.61331	.0029227
100's	503	1.08	.00587931	.0766767	7.12711	.00341884

Table 9

Estimation of Population Mean at 95% Confidence Limits

Tobacco and Cigarettes	Sample Mean $\bar{x}$	Population Mean $\bar{X} = \bar{x} \pm t(\frac{S_{\bar{x}}}{\sqrt{n}})$	Estimated Population Mean
	($)	($)	($)
Tobacco	1.41	$1.41 \pm 1.96(.0061299)$	$1.40 \leq \bar{X} \leq 1.42$
Cigarettes -			
All sizes	1.07	$1.07 \pm 1.96(.00166089)$	$1.06 \leq \bar{X} \leq 1.07$
Regular size	1.06	$1.06 \pm 1.96(.00229773)$	$1.05 \leq \bar{X} \leq 1.07$
King size	1.07	$1.07 \pm 1.96(.0029227)$	$1.06 \leq \bar{X} \leq 1.08$
100's	1.08	$1.08 \pm 1.96(.00341884)$	$1.07 \leq \bar{X} \leq 1.09$

$$\bar{X} = \bar{x} \pm t(\frac{S_{\bar{x}}}{\sqrt{n}})$$

Where t = 1.96 at 95% confidence probability.

Table 10

Estimation of Optimum Sample Sizes at Various Levels of Tolerated Errors

Tobacco and Cigarettes	Optimum Sample Size n_i Required									
	n_1 $.1	n_2 .09	n_3 .08	n_4 .07	n_5 .06	n_6 .05	n_7 .04	n_8 .03	n_9 .02	n_{10} .01
Tobacco	8	10	13	16	22	32	50	89	201	803
Cigarettes										
All sizes	2	2	3	4	5	7	11	20	45	179
Regular size	1	1	2	2	3	5	8	13	30	120
King size	2	2	3	4	5	8	12	21	48	193
100's	2	3	4	5	6	9	14	25	56	225

$$n_i = t^2 S^2 / e^2 = \left(\frac{tS}{e}\right)^2$$

Where n_i = Optimum Sample Size required

t = 1.96 at 95% Confidence Level

$S_{\bar{x}}^2$ = Variance and $S_{\bar{x}}$ = Standard Deviation

e = tolerated error in terms of dollar per unit

$.1, .09, .08, .07, .06, .05, .04, .03, .02, .01.

Table 11

Comparison of Statistical Variables and Parameters Between Metro Toronto and Outside Metro

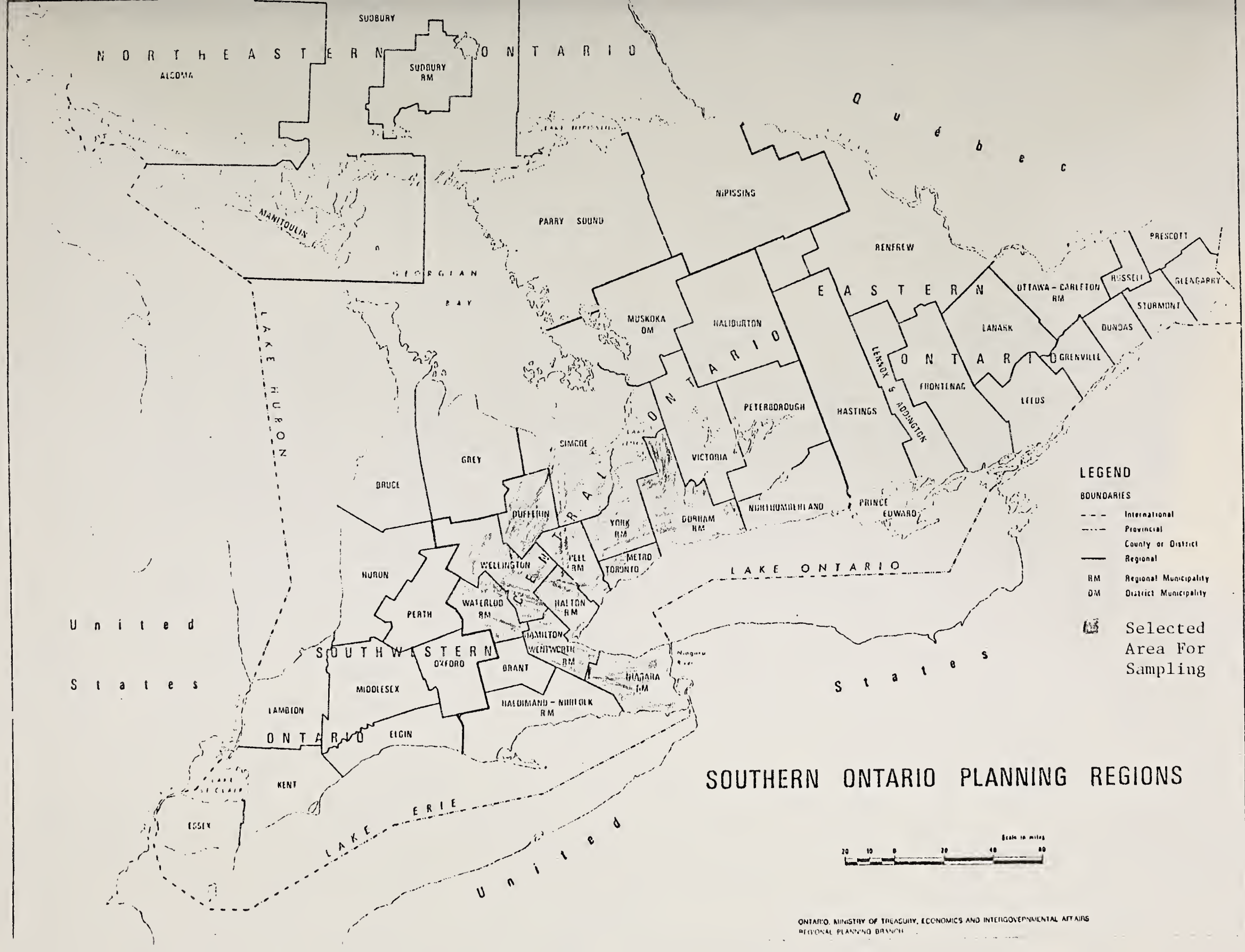
Survey Variable	Metro Toronto						Outside Metro					
	Sample	Mean	Pop.Mean Lower-Upper	Median	Mode	S.D.	Sample	Mean	Pop.Mean Lower-Upper	Median	Mode	S.D.
Fuel												
Regular Leaded	191	27.6	27.4 - 27.8	27.2	27.2	1.50861	266	27.8	27.6 - 28.0	27.2	27.2	1.39301
Regular Unleaded	189	29.9	29.6 - 30.2	29.3	29.2	1.78444	263	29.9	29.7 - 30.1	29.3	29.2	1.33955
Premium Leaded	15	30.5	29.5 - 31.5	29.4	29.4	1.94219	17	31.9	31.1 - 32.7	32.0	32.0	1.64079
Premium Unleaded	170	31.1	30.8 - 31.4	30.4	30.2	1.71564	209	31.1	30.9 - 31.3	30.6	30.2	1.56663
Diesel	48	26.5	26.1 - 26.9	26.3	28.0	1.57059	50	26.1	25.7 - 26.5	25.9	25.9	1.55308
Tobacco	400	1.42	1.41 - 1.43	1.43	1.12	0.143559	156	1.39	1.37 - 1.41	1.41	1.22	.146034
Cigarettes - All sizes	1,235	1.06	1.06 - 1.06	1.05	1.05	.0694583	449	1.09	1.08 - 1.10	1.08	1.05	.0597437
Cigarettes - Regular	427	1.05	1.05 - 1.06	1.05	1.05	.052636	165	1.09	1.08 - 1.10	1.08	1.10	.0558892
Cigarettes - King size	427	1.06	1.05 - 1.07	1.05	1.05	.0726232	162	1.09	1.08 - 1.10	1.09	1.10	.0616141
Cigarettes - 100's	381	1.07	1.06 - 1.08	1.05	1.05	.0802266	122	1.09	1.08 - 1.10	1.07	1.05	.0624961

Table 12

The Arithmetic Mean by Route Number and Grade of Fuel

Route Number	Diesel	Regular Leaded	Regular Unleaded	Premium Leaded	Premium Unleaded
Route 1	26.4¢	28.0¢	30.1¢	32.5¢	31.4¢
Route 2	25.9	28.0	30.1	31.3	31.3
Route 3	25.7	27.7	29.9	30.8	31.1
Route 4	25.8	27.0	29.1	29.0	30.6
Route 5	26.4	N/A	N/A	N/A	N/A
Route 6	26.8	28.3	30.0	N/A	31.7
Route 7	27.5	27.7	30.0	29.9	31.0
Sample Mean by Route	26.35	27.78	29.86	30.70	31.18
Sample Mean by Total	26.30	27.70	29.90	31.20	31.10
Route - Total	0.05¢	0.08¢	-0.04¢	-0.50¢	0.08¢

7. APPENDIX


NORTHEASTERN ONTARIO
SUDBURY
ALGOMA
SUDBURY RM
MANITOULIN
PARRY SOUND
NIPISSING
RENFREW
GEORGIAN BAY
LAKE HURON
MUSKOKA DM
HALIBURTON
EASTERN ONTARIO
PRESCOTT
RUSSELL
GLENGARRY
OTTAWA - CARLETON RM
STORMONT
DUNDAS
LANARK
GRENVILLE
LENNOX & ADDINGTON
FRONTENAC
LEEDS
PETERBOROUGH
HASTINGS
SIMCOE
GREY
BRUCE
DUFFERIN
YORK RM
DURHAM RM
VICTORIA
NORTHUMBERLAND
PRINCE EDWARD
CENTRAL ONTARIO
PEEL RM
METRO TORONTO
LAKE ONTARIO
NURON
WELLINGTON
WATERLOO RM
HALTON RM
PERTH
HAMILTON WENTWORTH RM
SOUTHWESTERN ONTARIO
OXFORD
BRANT
NIAGARA RM
MIDDLESEX
HALDIMAND - NORFOLK RM
LAMBTON
ELGIN
KENT
ESSEX
LAKE ERIE
United States
United States
LEGEND
BOUNDARIES
International
Provincial
County or District
Regional
RM Regional Municipality
DM District Municipality
Selected Area For Sampling
SOUTHERN ONTARIO PLANNING REGIONS
Scale in miles
20 10 0 20 40 60
ONTARIO. MINISTRY OF TREASURY, ECONOMICS AND INTERGOVERNMENTAL AFFAIRS
REGIONAL PLANNING BRANCH